TECHNICAL NOTES

General Technical Background to the 1996 Health Status Survey

Introduction

The purpose of this section is to provide the reader with a general methodological overview of the project. Persons interested in obtaining additional or more detailed information may contact:

Bureau of Surveillance and Analysis
Office of Public Health Data
Utah Department of Health
288 North 1460 West
Box 142875
Salt Lake City, UT 84114-2875
Phone: (801) 538-6108
E-mail: hlhda.phdata@state.ut.us

Sample Design

The 1996 Utah Health Status Survey represents the third such survey; previous surveys were conducted in 1986 and 1991. The statistical estimates in this report are based on *1996 Utah Health Status Survey* data.

The sample was a **complex survey sample** designed to be representative of all Utahns. It is best described as a weighted probability sample of approximately 6,300 households disproportionately stratified by twelve local health districts that cover the entire state. Five hundred household interviews were conducted in each health district, except Salt Lake City/County Health District, in which eight hundred household interviews were conducted in order to increase the precision of statewide estimates.

A **single stage, non-clustered, equal probability of selection telephone calling design** was used to generate telephone numbers, more specifically referred to as the *Casady-Lepkowski* (1993) calling design. This method begins by building a *base sampling frame* consisting of all possible telephone numbers from all working prefixes in Utah. Telephone numbers are arranged sequentially into groups of 100 by selecting all telephone numbers within an area code and prefix, plus the first and second digits of the suffix (e.g., 801-538-10XX represents a group that includes all 100 phone numbers between 801-538-1000 and 801-538-1099). Each group of 100 telephone numbers is classified as either high density (at least one residential listing) or low density (no listed residential phone numbers in the group). All low density groups are removed, and high density groups are retained. Telephone numbers are randomly selected from the high-density list. This sampling design ensures that both listed and unlisted phone numbers are included in the sample.

The survey interview was conducted with **one randomly-selected adult** (age 18 or older) in each household. To select this person, Gallup interviewers collected household membership information from the household contact person (the person who answered the telephone). One household member was then selected at random from the list of all household members age 18 or over. Survey questions were then asked about either, 1) all household members, 2) the survey respondent only, 3) a randomly selected adult or child household member (selected using the same method as was used to select the respondent), or 4) the household as a whole. Thus, the survey sample varies, depending on the within-household reference sample that was used for each set of survey questions. Each within-household reference sample has known probabilities of selection and can be generalized to the Utah population.

Survey Data Collection

The Utah Department of Health contracted with The Gallup Organization to collect the survey data. Gallup incorporated the telephone survey instrument into a **computer-assisted random digit dialing software program,** called SURVENT. Interviews were conducted by trained interviewers in a supervised environment across six sites. Interviews were conducted in Spanish when appropriate.

Computer-assisted telephone interviewing was chosen as the method of data collection for several reasons. First, it yields higher response rates, thus resulting in a more representative sample and reducing the amount of bias inherent in mail survey response rates. Second, it helps reduce non-sampling error by standardizing the data collection process. Data-entry errors are reduced because interviewers are not allowed to enter non-valid codes. It was also efficient because it allowed interviewers to enter responses directly into the database.

The survey questionnaire was divided into ***core*** and ***supplemental modules***. Core questions were asked of all households in the sample. Table 1 describes the types of "core" questions that were asked, and about whom they were asked. Notice that *not all questions were asked with regard to everyone in the household*.

Table 1.
CORE MODULE QUESTIONS

Question Topic	Within-Household Reference Sample
Demographic characteristics	All household members
Presence of chronic medical condition	All household members
Health insurance status	All household members
Injury incidence/safety issues	All household members
Lifestyle (smoking, drinking, exercise)	All household members
Subjective mental/physical health (SF12)	Respondent only (randomly-selected adult)
Health screening exam usage	Respondent only (randomly-selected adult)
Access to care/primary provider	Randomly-selected household member of any age
Household-level demographic characteristics	The household as a whole

In addition to the core survey questions (above), one of six different *supplemental modules* was administered to primarily non-overlapping randomly-assigned subsets of (approximately 1,000) households. Table 2 shows the types of questions asked in the supplemental module questions, and about whom they were asked.

Table 2.
SUPPLEMENTAL MODULE QUESTIONS

Type of Question	Within-Household Reference Sample
Limitations of activities	All household members
Migration	Respondent only (randomly-selected adult)
Health Plan Consumer Satisfaction	Respondent only (randomly-selected adult)
Fertility	Respondent or spouse only
Health Care Utilization	Randomly-selected household member of any age
Interpersonal violence	The household as a whole

*Note: All supplemental module questions were asked only of a subset of households.

While both core and supplemental modules yielded sufficient sample sizes to construct state-level estimates for the Utah population, the information collected from supplemental modules was not intended for use in district-level analyses.

Cooperation rate

The interview process took place over a three month period (from June to August, 1996), and resulted in a cooperation rate of 66.3%. If necessary, up to nine telephone attempts were made to contact a selected household. After a randomly-selected survey respondent was identified, up to nine attempts were made to conduct the interview with that person.

Weighting and Estimation Methods

Post-survey weighting adjustments were made so that the Health Status Survey findings could be more accurately generalized to Utah's population. Two types of post-survey weighting adjustments were made, one that adjusted for random sampling variation, and one that adjusted for disproportionate sampling (such as the over-sampling of smaller local health districts across the state). Although the two types of adjustment are distinct conceptually, they were accomplished in a single step.

The post-survey weighting adjustments weighted the sample to be proportionately consistent with the age, sex, geographic, and Hispanic status distribution of the 1996 Utah population. Utah population estimates by sex, single year of age, and county of residence were provided by the Utah Governor's Office of Planning and Budget (GOPB) (the estimates used were those compiled in 1994). Estimates of Utah's Hispanic population for 1996 were derived by calculating the average annual rate of increase of Hispanic persons for each health district using data from 1990 to 1994 Bureau of the Census reports, and then projecting those increases to 1996 GOPB local health district population counts. Total state estimates for Hispanic persons were calculated by summing across local health districts.

Separate post-survey weighting variables were constructed for use with each different subsample (e.g., a single local health district versus the entire state, respondents-only versus all household members, etc.). In all, there are 14 different weight variables that are used according to which questions are being analyzed and whether the user wants to generalize to a local health district or the entire state of Utah.

The post-survey weighting variables adjusted for the following factors:

1. The number of **phones** in the household.
2. The total **number of persons in the household** to which the data will be generalized (1 for questions that were asked about every household member, the number of adults in the household for questions that were asked only of the respondent, the number of persons in the household for questions that were asked of a randomly-selected household member).
3. The proportion of **Hispanic persons** in each local health district.
4. The **age and sex** distribution of each local health district.
5. The probabilities of selection for each **local health district**.

Population count estimates. Producing the population count estimates in the reference tables involved a number of steps. Once a percentage was calculated (e.g., the percentage uninsured) using appropriately weighted survey data, a population count (N) to which the percentage applied was estimated. In some cases analyses referenced certain age or sex groups, Hispanic persons or combinations of Utah counties. These total population group counts were readily available from the sources described earlier. However, for other groups where population counts were largely unavailable (e.g., analyses that examined the distribution of adult males by marital status), the population counts were estimated. This was achieved by multiplying the appropriate 1996 population total for that group (from 1996 GOPB estimates) by a proportion obtained from a frequency distribution or cross tabulation analysis of survey data. For instance, to calculate a population count for adult males who were married, the population of adult males from GOPB was multiplied by percentage of married adult males in the 1996 Utah Health Status Survey sample. Thus, any population count estimates not derived directly from existing age, sex, Hispanic status or county population estimates were derived from 1996 Health Status Survey data.

Missing Values. Another consideration that affected the presentation of the population estimates in table format was the inclusion or exclusion of missing values ("don't know" and "refused to answer"). Population percentage estimates were calculated after removing the "don't know" and "refused to answer" responses from the denominator. This, in effect, assumed that persons who gave these answers were distributed identically on the variable of interest to those who gave a valid answer to that variable. For instance, that among those who did not know whether they were insured, we assumed that 90.47% of them were insured and 9.53% were not insured -- percentages identical to those found among the sample members who answered the question with a valid response.

Removing the missing cases from an analysis is rather simple and straightforward for analyses of a single variable. However, when one variable is cross-tabulated by another variable, all missing cases from both variables must be removed from the analysis. Removing the missing cases in itself is not a problem. However, a problem is encountered when a population estimate for a given variable, such as the percentage of all Utahns that have health insurance, differs slightly from an analysis of "all Utahns" versus an analysis of "all Utahns by age group." This is because the missing cases on the age variable have been removed from one analysis and not from another. Since the percentage of all Utahns that have health insurance was calculated on slightly different samples, the result is slightly different. This problem was resolved by reporting the best population estimate available for any given population subgroup. For instance, in the table of insurance rates for all Utahns by age, the population estimate from an analysis that includes all Utahns, regardless of whether they reported missing values on the age variable has been substituted for the original total row in that table. The only drawback to this strategy is that the population count figures for Utahns with and without health insurance in tables like the

"Utahns by Age Group" table do not sum to the same number derived from the analysis of all Utahns regardless of whether they had missing values on the age variable. As a result, the tables appear as though they do not "add up."

Limitations and Other Special Considerations

Estimates developed from the sample may differ from the results of a complete census of all households in Utah due to two types of error, sampling and non-sampling error. Each type of error is present in estimates based on a survey sample. Good survey design and data collection techniques serve to minimize both sources of error.

Sampling error refers to random variation that occurs because only a subset of the entire population is sampled and used to estimate the finding in the entire population. It is often mis-termed "margin of error" in popular use, and is typically expressed as the "plus or minus" term, as in the following example:

> "The percentage of those polled who said they would vote for Bill Clinton was 52%, plus or minus 2%."

Because local health districts were disproportionately stratified and then weighted to reflect the Utah population, the sample was considered a complex survey sample design. Estimating the sampling error for a complex survey design requires special statistical techniques, derived from the standard error for each estimate. SUDAAN software (Research Triangle Institute) was chosen to estimate the standard errors of the survey estimates because it employs a statistical routine (Taylor-series expansion) that accounts for the complex survey design.

Reference tables in this report include estimates of sampling error expressed as 1.96 standard errors around (plus or minus) the estimate. As such, the estimates express the "95% confidence interval," or the interval that defines where the parameter would fall (with 95% probability) if all households in Utah were interviewed. In other words, there is only a 5% chance that the actual population parameter, or value, would fall outside the confidence interval. Figures in this report include bars showing this estimated variation around the parameter estimate. Readers should note that we have always presented the confidence interval as though it were symmetric, that is, of equal value both above and below (plus and minus) the estimate. It is often the case, however, that a confidence interval will be nonsymmetric. This occurs when the distribution is positively or negatively skewed, such as when a percentage is close to 0% or 100%. However, because the software program we use provides only symmetric confidence intervals, we are unable to provide the asymmetric estimates.

Non-sampling error also exists in survey estimates. Sources of non-sampling error include idiosyncratic interpretation of survey questions by respondents, variations in interviewer technique, household non-response to questions, coding errors, and so forth. No specific efforts were made to quantify the magnitude of non-sampling error.

Comparability with other surveys is an issue with all surveys. Differences in survey design, survey questions, estimation procedures, the socio-demographic and economic context, and changes in the structure and financing of the health care delivery system may all affect comparison between the 1996 Utah Health Status Survey and other surveys, including those conducted by the U.S. Bureau of the Census, the Behavioral Risk Factor Surveillance System surveys, and previous Utah Department of Health, Health Status Surveys.

Telephone surveys exclude certain population segments from the sampling frame, including persons in group living quarters (e.g., military barracks, nursing homes) and households without telephones. At the time of the 1990 Decennial Census, only four percent of Utah households were without telephone service. Typically, telephone surveys are biased because telephone households under-represent lower income and certain minority populations. In addition, studies have shown that non-telephone households tend to have lower rates of health care utilization (especially dental care), poorer health habits and health status, and lower rates of health insurance coverage (Thornberry and Massey, 1988).

Despite these overall disparities between telephone and non-telephone households, new survey research (Keeter, 1995) suggests that a similarity exists between data from non-telephone households and telephone households that experienced an interruption in service over the past 12 months. This similarity exists because many, if not most, households currently without telephones did have service in the recent past, and will have service again in the future. Therefore, certain households with telephones (those that had a recent interruption in service) are representative of "nonphone" households, allowing health status survey estimates that have been corrected for telephone noncoverage bias to be produced where indicated.

Analysis of the SF-12 Scale

Introduction

The purpose of this section is to provide a more thorough treatment of the methodology that was used to compute the SF-12 scales and difference scores used in this report. Readers interested in using the SF-12 items should register their intent with the Medical Outcomes Study group, and may be interested in obtaining technical and scoring manuals directly from them at the following address: The Health Institute, New England Medical Center Hospitals, Inc. Box 345, 750 Washington Street, Boston, MA 02111 (Ware et al., 1994, 1995).

This section is intended to provide only additional information that pertains specifically to the Utah administration of the SF-12 in the context of the Utah Health Status Survey. General information on the administration of the 1996 Utah Health Status Survey may be found in the section entitled General Technical Background to the 1996 Health Status Survey.

Brief Background of the SF-12

The SF-12 is a self-report measure of a person's perceived health on a number of dimensions (e.g., general health status, pain, depression, etc.). It was designed to measure patient outcomes in medical practice and clinical research for a variety of purposes, such as to monitor transitions in health status over time for diverse groups, to measure the burden of populations suffering from chronic medical and psychiatric conditions compared to well populations, to evaluate the relative benefits of different treatments, and to compare health outcomes across different health care delivery systems (McHorney et al., 1993, 1994). The Medical Outcomes Study group developed the SF-12 with the following objectives in mind 1) to serve as a measure of overall health status that took the patient's perspective into account, 2) to meet the need for a standardized health status measurement tool that was comprehensive, psychometrically sound, and brief (Ware & Sherbourne, 1992).

The SF-12 is the most recent in a series of health status measures developed by the Medical Outcomes Study group. Early on there were 18-item and 20-item measures. More recently, a 36-item short-form health status scale (SF-36) has replaced the earlier versions. The SF-36 can be scored to yield two overall measures: Physical health and mental health summary measures. Each measure is composed of eight subscales, representing eight different dimensions of physical and mental health:

Physical functioning,
Role functioning (physical),
Bodily pain,
General health,
Vitality,
Social functioning,
Role functioning (emotional), and
Mental health.

All eight subscales (and, hence, all 36 items) are used to form both the physical and mental health summary measures. The first four dimensions are weighted more heavily in the construction of the physical health summary score, while the second four dimensions are weighted more heavily in construction of the mental health summary score. The SF-36 can discriminate relatively well between persons with minor medical conditions, serious physical conditions, psychiatric conditions, and those with both serious physical and psychiatric conditions (McHorney, et al., 1993).

The SF-12 is not intended to replace the SF-36. Rather, a subset of 12 items was selected from the SF-36 because 36 items are often too many to include on a questionnaire (as was our experience with the Utah Health Status Survey). The 12-item subset explains over 90% of the statistical variance in the original 36-item physical and mental health summary scale measures, it can be scored so that it reproduces the average scores for the summary measures with a high degree of comparability, and it can be printed on one to two pages of a self-administered questionnaire or administered by an interviewer in less than two minutes, on average (Ware, Kosinski, & Keller, 1996).

Data Collection

The Utah Health Status Survey interview began with a set of questions on the general characteristics (e.g., age, height, weight, race) of each household member. One SF-12 item, (GH1, "In general, would you say your/[name's] health is poor, fair, good, very good, or excellent?") was positioned near the beginning of this series of items, and was asked with respect to each household member. The remaining SF-12 questions were administered immediately following the general demographic questions to avoid the context effects that other material in the survey (e.g., questions about chronic conditions and doctor visits) might have upon responses to the SF-12 questions. Aside from the general health item, the SF-12 questions were administered only to the survey respondents. The respondent was not asked to provide information on other persons in the household because it was believed that he or she could not provide accurate proxy data regarding the subjective states of other persons in the household. As a result, the SF-12 results reported here were derived from the responses of the 6,131 randomly-selected adult respondents, and are representative of persons age 18 and over in Utah.

Data Analysis

Initial Scoring. The SF-12 items were scored according to the procedure provided in Ware, et al. (1995). Initially, the 12 items are "dummy-coded" and weighted according to the SF-12 scoring manual. (Dummy-coding is a process that creates one variable for each item response. It is used to analyze ordinal-level data with parametric statistical techniques, such as linear regression.) This process was used to compute the Physical Health Composite Score and the Mental Health Composite Score summary measures of the SF-12. (A detailed description of this process may be obtained from the Bureau of Surveillance and Analysis, Utah Department of Health.)

The weighting algorithm was designed so that the SF-12 scores are consistent with the SF-36 scores, that is, each has a national mean of 50 and standard deviation of 10. Higher scores in both physical and mental health measures indicate better health. Scores higher than the mean indicate

that a person has better health status than average, while scores lower than the mean indicate poorer health status than average.

In Utah, as in national samples of the general population, the distribution was quite negatively skewed, with a range of approximately 10 to 70. Given this distribution of scores, persons with poor health outcomes could score much lower than than the mean, as many as 40 points lower, but persons with excellent health outcomes could score only as many as 20 points above the mean.

Age-Specific Difference Scores. The physical and mental health summary measures differ by age group, with older persons experiencing worse physical health, but better mental health outcomes than their younger counterparts.

Summary Measure Scores by Age Group

Age Group	Physical Health	Mental Health
18-34	51.93	52.21
35-44	51.42	52.35
45-54	50.53	53.64
55-64	47.67	54.69
65-74	44.59	56.50
75+	42.71	55.03

Because of this pattern of responding, the Medical Outcomes Study group recommends that a person's score be interpreted in the context of his or her own age group. Because we wanted to compare across various population groups while controlling for the effects of age, we wanted to form a single score that would take into account age differences in responding.

We had also heard from local users of the SF-36 and SF-12 that the scales were not intuitive. According to psychometric scaling theory, a scale is a much more powerful measurement tool than a single item. Single items are prone to error, such as differences in interpretation by respondents. A scale is also advantageous because it can measure more of the richness of a phenomenon, such as measuring all eight dimensions of health status, ensuring that the full range of experiences is represented in the data. However, scales also have a disadvantage: They are often less intuitive than a single item. For instance, the general health item of the SF-12 elicits a person's health status as either excellent, very good, good, fair, or poor. It is easy to imagine what is meant by excellent health versus poor health. One can not look at a person's SF-12 score and know what it means. One user reported, "So I learn that a person has a score of 42.5. What does that tell me?"

Age-specific difference scores were derived in response to both the need for a single score that controlled for the effects of age, and the need for a scale that is more intuitive. The age-specific difference score is the difference between a person's score and his or her age-specific reference group. Thus, if a person has a difference score of -5.5, it indicates that they scored 5.5 points lower than other persons their age — an indication of somewhat poor health. The difference score

is perhaps the most intuitive way to judge a persons score. Looking at a difference score, it is immediately clear whether a person is more healthy or less healthy than other persons in their comparison group. In addition, difference scores can be compared across age groups. That is, a score of -5.5 means virtually the same thing, regardless of a person's age.

Developing Cut-Points for Above- and Below-Average Scores. After computing age-specific difference scores, the SF-12 scales were more intuitive than they had originally been. Positive scores indicated good health, whereas negative scores indicated poor health. But there was still a question of how low a person's score had to be in order for him or her to be considered in poor health. The difference score indicated the direction and magnitude of the score, but it did not indicate at which point a negative score should elicit concern.

The standard error of measurement is used to assign these cut-off points to individual scores. The SEM is a psychometric property of the scale that indicates the extent to which an individual's score should be expected to vary over a large number of randomly parallel tests (given that his or her health has not changed) (Kosinsky, 1997; Ware, Bayliss, Rogers, Kosinski & Tarlov, 1996; Nunnally, 1978). It is computed as follows:

SEM=std. dev. * (sqrt (1-reliability coefficient))

The unweighted sample data were used to compute the reliability coefficient (also known as Cronbach's alpha). Weighting the survey sample was deemed unnecessary for this step because the reliability coefficient is a property of the scale that is based on the intercorrelation of items — we were not producing an estimate of a population parameter that would be generalized to the state population. In practice, weighting the data made very little difference in the value of the reliability coefficient (unweighted = .868, weighted = .853). The same reliability coefficient was used to compute the SEM for both physical and mental health summary measures because all 12 items are used in the computation of both scales.

Weighted sample data were used to calculate the standard deviations for the two scales. SUDAAN (a statistical package that uses Taylor series expansion to derive unbiased estimates of sampling variation) was not used to calculate the standard deviation. It was deemed unnecessary, since an estimate of the population parameter, standard deviation, was desired rather than an estimate of the sampling variation of the mean scale score (Williams, 1997). Standard deviations for the two scales were 9.16 and 8.56 for the physical and mental health summary scales, respectively.

The standard errors of measurement for the physical and mental health scales were multiplied by 1.96 to derive the 95% confidence interval, the theoretical range within which an individual's score would vary over 95% of a large number of repeated observations with parallel forms of the same test. Conceptually, this confidence interval should be applied to each individual's score to ascertain whether the individual's confidence interval includes the mean scale score. If their confidence interval includes the average score, then they should be considered "no different from average." In practice, however, the confidence interval may also be applied to the mean scale score to define a range, within which an individual score could be considered average. The 95% confidence intervals for the physical and mental health summary scores were 6.53 and 6.11, respectively.

BIBLIOGRAPHY

Casady, Robert J., and James M. Lepkowski (1993) Stratified telephone survey designs. Survey Methodology, 19 (1) 103-113.

Institute of Medicine (1988) The Future of Public Health. Washington D.C.: National Academy of Sciences, National Academy Press.

Keeter, Scott (1995) Estimating telephone noncoverage bias with a telephone survey. Public Opintion Quarterly, 59, 196-217.

Kosinsky, Mark, The Health Institute, New England Medical Center, personal communications, March 1997.

McHorney, Colleen A., John E. Ware, Jr., and Anastasia E. Raczek (1993) The MOS 36-item short-form health survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. Medical Care, 31 (3), 247-263.

McHorney, Colleen A., John E. Ware, Jr., J.F. Rachel Lu, and Cathy Donald Sherbourne (1994) The MOS 36-item short-form health survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. Medical Care, 32 (1), 40-66.

Nunnally, Jum C. (1978) Psychometric Theory. New York: McGraw-Hill Book Company.

Thornberry, Owen T. and James T. Massey (1988) Trends in United States telephone coverage across time and subgroups. In Robert M. Groves, Paul P. Biemer, Lar E. Lyberg, James T. Massey, William L. Nichols II, and Joseph Waksberg (eds.) Telephone Survey Methodology. New York: John Wiley and Sons.

U.S. Department of Health and Human Services, Public Health Service. (1991) Health People 2000: National Health Promotion and Disease Prevention Objectives. Washington, D.C.: U.S. Government Printing Office, publication # (PHS) 91-50212.

Ware, John E., Jr., Martha S. Bayliss, William H. Rogers, Mark Kosinski, and Alvin Tarlov (1996) Differences in 4-year health outcomes for elderly and poor, chronically ill patients treated in HMO and fee-for-service systems. Journal of the American Medical Association, 276 (13), 1039-1047.

Ware, John E., Jr., Mark Kosinski, and Susan D. Keller (1994) SF-36 Physical and Mental Health Summary Scales: A User's Manual. Boston, MA: The Health Institute, New England Medical Center.

Ware, John E., Jr., Mark Kosinski, and Susan D. Keller (1995) SF-12: How to Score the SF-12 Physical and Mental Health Summary Scales. Boston, MA: The Health Institute, New England Medical Center.

Ware, John E., Jr., Mark Kosinski, and Susan D. Keller (1996) A 12-item short-form health survey: Construction of scales and preliminary tests of reliability and Validity. Medical Care, 34 (3), 220-233.

Ware, John E., Jr., and Cathy Donald Sherbourne (1992) The MOS 36-item short-form health survey (SF-36). Medical Care, 30 (6), 473-481.

Williams, Rick, Research Triangle Institute, personal communication, March 1997.